

token sequence a meaning 22 that is relevant to a specific user application. In a typical implementation, language analyzer 14 compares a meaning token sequence with a set of semantic rules 24 that are defined by a grammar compiler. Language analyzer 14 identifies permissible meanings based upon the semantic rules, and
5 outputs summary structures representing permissible sentences. Application command translator 16 matches the summary structures with one of a set of application-specific actions 26, and carries out the selected action by issuing one or more commands 28 to a downstream user application for processing.

As explained in detail below, speech recognizer 12 may be a conventional
10 automatic speech recognition system that operates with respect to the meaning token dictionary in the same way that it would operate with respect to a conventional speech recognition dictionary. The output from the speech recognizer 12, however, is in the form of meaning tokens rather than simple transcriptions of ordinary spoken words. The meaning token dictionary may simplify the work that must be performed
15 by language analyzer 14 and any downstream application program that uses an automatic speech recognition system for spoken language input. In addition, the output from speech recognizer 12 may be substantially free of ambiguities, allowing language analyzer 14 and the downstream application program to use deterministic interpretation algorithms (which are simpler and more efficient than non-
20 deterministic algorithms).

Referring to FIG. 2, in one embodiment, automatic speech recognition system
10 may be implemented as one or more respective software modules operating on a general-purpose computer 30. Computer 30 includes a processing unit 34, a system memory 36, and a system bus 38 that couples processing unit 34 to the various
25 components of computer 30. Processing unit 34 may include one or more processors, each of which may be in the form of any one of various commercially available processors. System memory 36 includes a read only memory (ROM) 40 that stores a basic input/output system (BIOS) containing start-up routines for computer 30, and a random access memory (RAM) 42. System bus 38 may be a
30 memory bus, a peripheral bus or a local bus, and may be compatible with any of a variety of bus protocols, including PCI, VESA, Microchannel, ISA, and EISA.

Computer 30 also includes a hard drive 44, a floppy drive 46, and CD ROM drive 48 that are connected to system bus 38 by respective interfaces 50, 52, 54. Hard drive 44, floppy drive 46, and CD ROM drive 48 contain respective computer-readable media disks 56, 58, 60 that provide non-volatile or persistent storage for data, data structures and computer-executable instructions. Other computer-readable storage devices (e.g., magnetic tape drives, flash memory devices, and digital video disks) also may be used with computer 30. A user may interact (e.g., enter commands or data) with computer 40 using a keyboard 62 and a mouse 64. A user also may interact with automatic speech recognition system 10, which is executing on computer 30, by speaking into a microphone 66. Computer 30 may output synthesized speech and other sounds through a speaker 68. Information may be displayed to the user on a monitor 70. Computer 30 also may include peripheral output devices, such as a printer. One or more remote computers 72 may be connected to computer 30 over a local area network (LAN) 74, and one or more remote computers 76 may be connected to computer 30 over a wide area network (WAN) 78 (e.g., the Internet).

Referring to FIG. 3, in one embodiment, speech recognizer 12 includes a spectrum analyzer 80, a pattern matching system 82, and a meaning token selecting system 84. Spectrum analyzer 80 operates on the digitized speech samples that are received from the acoustic input device to compute a sequence of spectrum frames 86. Pattern matching system 82 establishes a correspondence between the sequence of spectrum frames 86 and pre-stored (trained) representations of speech sounds 88, and produces a sequence of speech units 90. Meaning token selecting system 84 assembles the speech units 90 into sets of possible pronunciations and selects from a meaning token dictionary 92 a sequence of recognized meaning tokens 20 corresponding to a sequence of vocabulary words mostly likely to have been spoken by a user.

As shown in FIGS. 4 and 5, meaning token dictionary 92 maps a pronunciation 96 to a meaning token 20 having a dictionary spelling that signifies a single meaning (or thought or idea). As used herein the term "dictionary" refers broadly to a data structure that is stored on a computer-readable physical medium

and contains vocabulary words that may be mapped onto a spoken utterance on the basis of acoustic characteristics and, possibly, a-priori probabilities or another rule, such as a linguistic model or a grammar model. Meaning token 20 may have the same pronunciation as a vocabulary word in a conventional speech recognition dictionary, but the dictionary spelling that is associated with that vocabulary word is selected to signify a single meaning rather than a simple transcription of the words that were spoken. For example, the pronunciation of “do you have” could be the pronunciation for a meaning token that is spelled “seek”. The same meaning token also may be associated with many different pronunciations. For example, “seek” could also be the dictionary spelling for the pronunciations of “I want”, “I would like”, “I’m looking for”, “could you please give me”, and “gimme”.

Referring to FIG. 6, in some embodiments, the spelling of a meaning token may encode various kinds of information that may be used by language analyzer 14 or a downstream application program. In some embodiments, a meaning token may encode one or more labels identifying one or more respective application-specific categories, such as an object category, a place category, an event category, or an action category. For example, book titles may appear in meaning token dictionary 92 for a meaning token “book”, which may include the title or some other identifier. In the example of FIG. 6, the pronunciation “profiles in courage” could be spelled “book”, “book = ProfilesInCourage”, “book 12345”, or “bookISBN1579120148”. Similarly, the spelling of the “seek” meaning token could be “action = seek”. In this case, if a user says to the application “Do you have Profiles in Courage?”, speech recognizer 12 might output the two-token sequence “action = seek book = ProfilesInCourage,” rather than the ambiguous six word sequence “do you have profiles in courage”.

In some embodiments, meaning tokens may have longer spellings than the corresponding vocabulary words that are contained in a conventional speech recognition dictionary. Thus, each vocabulary word that is contained in the meaning token dictionary may have a unique spelling. In addition, multiple meaning tokens may be associated with each of one or more polysemous vocabulary words that are contained in the meaning token dictionary. In this way, language analyzer 14 and